レター Letter

r-compatibility of partial words

Tetsuo Moriya*

Abstract: In this paper, we study partial words in relation with pcodes, r-compatibility, and containment.

First, we introduce the concept r-compatibility, which is so called concept of compatibility of duality. We denote $C'(L)$, the set of all partial words r-compatible with elements of the set L.

In ["A note on pcodes of partial words," IEICE Trans. Inf. and Syst., vol.E97-D, no.1, January 2014], we introduced $C_{\subset}(L)$, the set of all partial words contained by elements of L, and $C_{\supset}(L)$, the set of all partial words containing elements of L, for a set L of partial words. We discuss the relation between $C'(L)$, $C_{\subset}(L)$, and $C_{\supset}(L)$.

Next, we consider the condition for $C'(L)$ to be a pcode when L is a pcode.

Key words: partial word, containment, compatibility, pcode

1. Introduction

Partial words are strings over a finite alphabet that may contain a number of "do not know" symbols. The motivation behind the notion of partial words is the comparison of two genes (or two proteins). Alignment of two such strings can be viewed as a construction of two partial words that are said to be compatible in a sense that will be described in Section 2

Codes play an important role in the study of combinatorics on words ([1], [9]). In [4], pcodes were introduced in relation with combinatorics on partial words. While a code L of words does not allow two distinct decipherings of some word in $L+$, a pcode K of partial words does not allow two distinct "compatible" decipherings in $K+$.

Some combinatorial properties of partial words have been investigated in previous studies ([2], [3], [4], [5], [7], [8], [10]).

In this paper, we study partial words in relation with r-compatibility, pcodes and containment. Let L be a set of partial words. In [6], the set $C(L)$ of all partial words compatible with the elements of a set L of partial words was defined.

In [11], we introduced the following two sets of partial words in relation with $C(L)$.

(1) $C_{\subset}(L)$, the set of all partial words containing elements of L, and

(2) $C_{\supset}(L)$, the set of all partial words contained by elements of L.

First, we discuss the relation between $C'(L)$, $C_{\subset}(L)$, and $C_{\supset}(L)$. Next, we consider the condition for $C'(L)$ to be a pcode when L is a pcode.

* School of Science and Engineering, Kokushikan University

2. Preliminaries

Let Σ be a nonempty finite set of symbols, which we call an alphabet. A word over the alphabet Σ is a finite sequence of elements of Σ. The empty sequence is called an *empty word* and is denoted by ε. The set of all words over Σ is denoted by Σ^*. The set of nonempty words over Σ is denoted by Σ^+. Thus, $\Sigma^+ = \Sigma^* \backslash \{\varepsilon\}$.

For w in Σ^*, $|w|$ denotes the length of w. A *language* over Σ is a set $L \subseteq \Sigma^*$.

A word of length n over Σ can be defined by a total function $u : \{0, 1, ..., n-1\} \to \Sigma$ and is usually represented as $u = a_0 a_1 ... a_{n-1}$ with $a_i \in \Sigma$.

A partial word (pword for short) of length n over Σ is a partial function $u : \{0, 1, ..., n-1\} \to \Sigma$. For $0 \le i < n$, if u is defined, then we say that i belongs to the domain of u (denoted by $i \in D(u)$); otherwise, we say that i belongs to the set of holes of u (denoted by $i \in H(u)$). A word over Σ is a partial word over Σ with an empty set of holes (we refer to words as *full words*). For any partial word u over Σ, $|u|$ denotes its length. Clearly, $|\varepsilon| = 0$. Let $W_0(\Sigma)$ denote the set Σ^*, and for $i \ge 1$, let $W_i(\Sigma)$ denote the set of partial words over Σ with at most i holes. We put $W(\Sigma) = \cup_{i \ge 1} W_i(\Sigma)$, the set of all partial words over Σ with an arbitrary number of holes.

If u is a partial word of length n over Σ, then the companion of u (denoted by $u_\diamond$) is the total function $u_\diamond : \{0, 1, ..., n-1\} \to \Sigma \cup \{\diamond\}$ dened as

$u_\diamond = u(i)$ if $i \in D(u)$, $\diamond$ otherwise.

The symbol $\diamond \notin \Sigma$ is considered the "do not know" symbol. The word $u = ab\diamond ab\diamond a$ is the companion of the partial word u of length 7, where $D(u) = \{0, 1, 3, 4, 6\}$ and $H(u) = \{2, 5\}$. The bijectivity of the map $u \mapsto u_\diamond$ allows us to define partial words concepts such as concatenation and powers, in a trivial manner. The set $W(\Sigma)$ is a monoid under the concatenation of partial words (ε serves an identity). For convenience in

the sequel, we say, for instance, *"the partial word $ab\diamond ab\diamond a$"* instead of *"the partial word with companion $ab\diamond ab\diamond a$"*.

Given two subsets L, K of $W(\Sigma)$, we define $LK = \{uv | u \in L \text{ and } v \in K\}$. We sometimes write $L \sqsubset K$ if $L \subset K$ but $L \neq K$.

A factorization of a partial word u is any sequence $u_1, u_2, ..., u_i$ of partial words such that $u = u_1u_2...u_i$. For a subset L of $W(\Sigma)$ and integer $i \geq 0$, let L^i denote the set $\{u_1u_2...u_i | u_1, ..., u_i \in L\}$. For a subset L of $W(\Sigma)$, we use the notation $\|L\|$ for the cardinality of L.

Let L^* denote the submonoid of $W(\Sigma)$ generated by L, or $L^* = \cup_{i\geq 0} L^i$, where $L^0 = \{\varepsilon\}$, and let L^+ denote the subsemigroup of $W(\Sigma)$ generated by L, or $L^+ = \cup_{i>0} L^i$. An element of $\{\diamond\}^+$ is called a *hole word*. If u and v are partial words of equal length, then u is said to be *contained* in v, denoted by $u \subset v$ or $v \supset u$ if all elements in $D(u)$ are in $D(v)$ and $u(i) = v(i)$ for all $i \in D(u)$. We sometimes write $u \sqsubset v$ if $u \subset v$ but $u \neq v$. The partial words u and v are compatible, denoted by $u \uparrow v$ if there exists a partial word w such that $u \subset w$ and $v \subset w$. Let $u \vee v$ denote the least upper bound of u and v. The partial words u and v are r-compatible, denoted by $u \downarrow v$ if there exists a partial word w such that $w \subset u$ and $w \subset v$ with $w \notin \{\diamond\}^+$. Let $u \wedge v$ denote the greatest lower bound of u and v.

Let $L \subseteq W(\Sigma)$. We dene $C(L)$, $C_\subset(L)$, $C_\supset(L)$, and $C'(L)$ as follows:

$C(L) = \{y \in W(\Sigma) | x \uparrow y \text{ for some } x \in L\}$.

$C\subset(L) = \{y \in W(\Sigma) | x \subset y \text{ for some } x \in L\}$.

$C\supset(L) = \{y \in W(\Sigma) | x \supset y \text{ for some } x \in L\}$.

$C'(L) = \{y \in W(\Sigma) | x \downarrow y \text{ for some } x \in L\}$.

Let L be a nonempty subset of $W(\Sigma)\backslash\{\varepsilon\}$. Then, L is a *pcode* if for all integers $m \geq 1$, $n \geq 1$ and partial words $u_1, ..., u_m, v_1, ..., v_n \in L$, the condition

$u_1...u_m \uparrow v_1...v_n$

implies that $m = n$ and $u_i = v_i$ for $i = 1, ..., m$.

Remark 1 *Containment on pwords is a partial order.*

The relation is trivially reflexive. The relation is anti-symmetric. To see this, suppose that $u \subset v$ and $v \subset u$. Then $D(u) = D(v)$ and $u(i) = v(i)$ for all $i \in D(u)$. Thus $u = v$. It is obvious that containment is transitive.

Remark 2 *Compatibility on pwords is an equivalence relation.*

Compatibility on partial words is trivially reflexive and symmetric. It is also transitive. To see this, suppose that $u \uparrow v$ and $v \uparrow w$. There exist a pword x and y such that $u \subset x$, $v \subset x$, $v \subset y$, and $w \subset y$. Let the least upper bound $x \vee y$ be z. Then $u \subset z$, $v \subset z$, and $w \subset z$. Hence $u \uparrow w$.

Remark 3 *r-compatibility on pwords is not equivalence relation. r-compatibility on pwords is trivially reflexive and symmetric. However, it is not transitive. For example, $a\diamond\downarrow ab$ and $ab\downarrow\diamond b$, but $a\diamond\downarrow\diamond b$ does not hold.*

3. r-compatibility of partial words

Proposition 1 *For $L \subseteq W(\Sigma)$, $C'(L) = C_\subset(C'_\supset(L))$, where $C'_\supset(L) = C_\supset(L)\backslash\{\diamond\}^+$*

Proof. Let $y \in C'(L)$. There exists $x \in L$ such that $x \downarrow y$, that is, there exists $z \in W(\Sigma)\backslash\{\diamond\}^+$ such that $z \subset x$ and $z \subset y$. It follows that $z \in C'_\supset(L)$ and that $y \in C_\subset(z) \subseteq C_\subset(C'_\supset(L))$. Thus, $C'(L) \subseteq C_\subset(C'_\supset(L))$.

Conversely, let $z \in C_\subset(C'_\supset(L))$. There exist $x \in L$ and $y \in W(\Sigma)\backslash\{\diamond\}^+$ such that $y \subset x$ and $y \subset z$. We have $x \downarrow z$. Hence, $z \in C'(L)$. Thus, $C_\subset(C'_\supset(L)) \subseteq C(L)$. ::

Next, we consider the condition for $C'(L)$ to be a pcode when L is a pcode.

Proposition 2 *Let $L \subseteq W(\Sigma)\backslash\{\varepsilon\}$.*

$C'(L)$ is a pcode iff $L \subseteq \Sigma^*$ and L is a pcode.

Proof.

[If] If $L \subseteq \Sigma^*$, then $C'(L)=L$. Thus, the result holds.

[Only if] Suppose that $L \not\subset \Sigma^*$. Then, there exists $x \in L$ such that $\|H(x)\| \geq 1$.

Moreover, there exists $y \in W(\Sigma)$ such that $y \in C'(L)$, $x \sqsubset y$, and $x \downarrow y$. We have also $x \uparrow y$. Since $x \in C_\subset(L)$, it follows that $C_\subset(L)$ is not a pcode.

Next, suppose that L is not a pcode. Since $L \subseteq C'(L)$, $C'(L)$ is not a pcode. ::

References

[1] J. Berstel and D. Perrin, Theory of Codes, Academic Press, New York, 1985.

[2] J. Berstel and L. Boasson, "Partial words and a theorem of Fine and Wilf," Theoretical Computer Science vol.218, pp.135-141, 1999.

[3] F. Blanchet-Sadri, "Periodicity on partial words," Computers and Mathematics with Applications vol.47, pp.71-82, 2004.

[4] F. Blanchet-Sadri, "Codes, orderings, and partial words," Theoretical Com-puter Science, vol.329, pp.177-202, 2004.

[5] F. Blanchet-Sadri and Ajay Chriscoe, "Local periods and binary partial words: an algorithm," Theoretical Computer Science vol.314, pp.189-216, 2004. http://www.uncg.edu/mat/AlgBin/.

[6] F. Blanchet-Sadri and M. Mooreeld. "Pcodes of partial words," Preprint. 2005. http://www.uncg.edu/cmp/research/pcode/.

[7] F. Blanchet-Sadri and S. Duncan, "Partial words and the critical factorization theorem," Journal of Combinatorial Theory, Series A 109, pp.221-245, 2005. http://www.uncg.edu/mat/cft/.

[8] F. Blanchet-Sadri and Robert A. Hegstrom, "Partial words and a theorem of Fine and Wilf revisited," Theoretical Computer Science vol.270, pp.401-419, 2002.

[9] H. J. Shyr, Free monoids and languages, Hon Min Book Company, Taichung, Taiwan, 2001.

[10] F. Blanchet-Sadri and D.K. Luhmann, "Conjugacy on partial words," Theoret-ical Computer Science vol.289, pp.297-312, 2002.

[11] T. Moriya and I. Kataoka, "A note on pcodes of partial words," IEICE Trans. Inf. and Syst., vol.E97-D, no.1, January 2014.